The participation of IntermidiaLab at the ImageCLEF 2012 Photo Annotation Task

Marcelo G. Manzato

Mathematics and Computing Institute, University of São Paulo
Av. Trabalhador Sancarlense, 400, PO Box 668 – 13560-970
São Carlos, SP – Brazil
mmanzato@icmc.usp.br

Abstract. This paper presents the results of our concept annotation tool in the ImageCLEF 2012 Photo Annotation Task. Our approach is based only on textual features, in particular collaborative user tags. It faces some of the challenges of using user-generated terms, such as noise and incompleteness. The first one is supported by a multi-phase filtering procedure that amends misspelled tags and considers only those semantically related to the actual context of the image. The second one, in turn, is reduced with a tag enrichment activity which aggregates related terms to the actual list of tags in order to make annotation more effective.

Keywords: Automatic annotation, concept classification, user tags, folksonomy, enrichment.

1 Introduction

The overload of multimedia content available on the Web has introduced new challenges for search engines to retrieve, organize and classify the data according to the user's needs. One important and common step of these tasks is the extraction of meaningful information describing the content. Typically, the provision of such metadata is accomplished by content authors or providers, who create well-structured information about specific media items, helping search and analysis tasks. Automatic approaches may also be adopted, which create multimedia descriptions without human intervention, analyzing low level visual features in order to infer semantic information.

It is known that automatic and manual indexing techniques have limitations that may affect the process of gathering semantic information. Indeed, the lack of methods to provide meaningful descriptions has been named semantic gap [15], and over ten years researchers have been working on the development of strategies to overcome related problems. For instance, in case of automatic approaches, usually they are highly dependent on the content domain, once they infer semantic information based on low-level visual features [2]. In manual approaches, in turn, the description is considered a time-consuming and error prone task [13].

In face of the huge amount of multimedia indexing techniques proposed so far, the ImageCLEF competitions[1] play an important role in the context of image retrieval. Each year, a set of specific challenging problems is proposed to researchers with the objective to push forward the state-of-art, allowing the resulting techniques to be discussed and compared against each other using a unique and standardized set of requirements. One of these problems is the Photo Annotation Task, which consists of a multi-label classification challenge, that aims to analyze a collection of Flickr[2] photos in terms of their visual and/or textual features in order to detect the presence of one or more concepts.

Particularly in the case of textual features, the dataset of images made available by ImageCLEF to be annotated includes Flickr tags, which are a set of terms created by users to facilitate further search by themselves. Considering other benefits, some authors [9][20][1][21] have explored collaborative tagging to create folksonomies and semantic relationships based on co-occurrence of tags. With such structures, it is possible to gather semantic information from the content in a collaborative way, reducing the main problems related to traditional indexing approaches.

This paper presents the techniques developed by our group to handle the ImageCLEF 2012 Photo Annotation Task [18]. Unlike the majority of the participants who explore visual features or a combination of visual and textual features, we focus only on user tags. Although many of the defined concepts are difficult to be found using only text, we believe exploring textual information is a good start point to design a robust and multimodal annotation approach.

This paper is structured in the following way. Section 2 presents the related work which depict other text-based approaches proposed in previous years of ImageCLEF. Section 3 provides the prior steps in our algorithm which consist of a preparation procedure to gather semantic information from the training set. Section 4 describes the concept annotator algorithm proposed in this paper. Section 5 presents the official results of the evaluation. Finally, Sections 6 and 7 present the final remarks, future work and acknowledgments.

2 Related Work

Analyzing the last two years of ImageCLEF Photo Annotation Task, it can be noted that the majority of approaches is based on visual features or a combination of visual and textual features. In 2010, from 54 groups registered for the task, only two groups submitted runs in the textual category only. One example is the work of Li et al. [9], which used textual features (user tags and EXIF information) to perform automatic annotation. Their approach starts with a document expansion procedure which assigns additional content to concepts and image tags by consulting external information resources, such as DBpedia[3]. After that, expanded textual metadata is compared to expanded concepts in

[1] http://imageclef.org/
[2] http://www.flickr.com/
[3] http://dbpedia.org

order to make concept assumptions. They also provide a method to assign concepts to images by analyzing EXIF data, such as the time a photo was shot. Finally, additional concepts are considered by infering affiliations and opposite relations among them.

In 2011, 48 groups registered for the challenge, and seven submitted at least one run in the textual category only. One of the submissions of Daróczy et al. [6], for instance, was a text-based approach that uses Flickr tags to create a kernel weighting procedure based on similarity of images which is computed using the Jensen-Shannon divergence.

In addition to this similarity calculation, other metrics were also used, such as the ones based on WordNet[4] ontology. Indeed, Liu et al. [10] proposed a method to extract text features for concept detection based on a semantic distance between words, which are expected to capture the semantic meanings from images. In total, ten features were extracted, which are based on different types of information, including semantic similarities according to WordNet ontology, affective norms for English words, etc.

Znaidia et al. [21] submitted a text-based approach that uses two distances among tags and visual concepts: one based on Wordnet ontology and the other based on social networks. The correlation between tags and visual concepts is also explored by Amel et al. [1], who proposed an approach that uses Flickr tags to extract contextual relationships of tags and concepts. Two types of contextual graphs are modeled: an inter-concepts graph and a concept-tags graph. The similarity among tags and concepts is computed based on the principle of Google distance [5].

In order to use the Flickr tags, it is important to pre-process them to reduce the occurrence of noise, and give more importance to particular terms. These two steps are performed by Izawa et al. [7], who submitted an approach that first performs a morphological analysis on all terms of the training images. Next, a tf-idf score is assigned to each tag. Then, the system stores the correspondence information for the tags and concepts in a casebase. Finally, the annotation on a given test image is accomplished by matching the tags attached to the image to the tags of the casebase.

The tf-idf score is also used by Nagel et al. [12], who submitted a text-only approach consisting of a pre-processing step on Flickr tags to reduce noise, followed by tf-idf assignment for each term in order to construct a text-based descriptor.

Spyromitros-Xioufis et al. [16] proposed a textual model that uses the boolean bag-of-words representation, in addition to stemming, stop words removal, and feature selection using the chi-squared-max method [8]. After that, the extracted features were used in a multi-label learning algorithm based on Ensemble of Classifier Chains [14]

The aforementioned text-based approaches differ from ours because we provide a co-occurrence-based tag enrichment phase in order to overcome the incompleteness of terms in resources that do not have enough associated tags. In

[4] http://wordnet.princeton.edu

addition, based on the enriched tags list, we also calculate the co-occurrence of concepts and tags during training to highlight the most frequent concepts by given tags.

3 Training

Our automatic annotator starts with a training procedure whose main objective is to gather semantic information from the available corpus already annotated. In addition, all user tags must be filtered and analyzed in order to remove incomplete and/or incorrect tags. Figure 1 illustrates the overall schema. Although in this participation we are annotating only images, our approach may be used with any other multimedia modality, such as video and audio, unless the content does not have any associated tags.

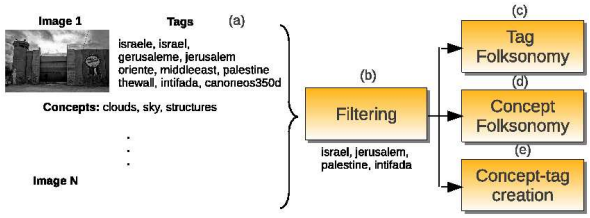

Fig. 1. Training procedure.

All available tags about an image (Figure 1 (a)) are submitted into a module which will filter incorrect and/or incomplete terms. This procedure has the objective to eliminate noise which can affect later the quality of concept classification. After that, all terms are used to create a folksonomy of tags (Figure 1 (c)), so that semantically related terms can be obtained during the annotation of images with few associated tags. A folksonomy of concepts is also constructed based on their co-occurrence, as indicated in Figure 1 (d). The idea behind this process is to identify semantically related concepts, and then, assign those related concepts based on the ones already found. For instance, if an image was annotated with the concept "combustion_flames", so it will also be annotated with "combustion_smoke". Finally, we use both concepts and terms of each image to create a table containing the most frequent concept for each tag (Figure 1 (e)). This table will be used during classification to predict possible concepts according to tags assigned by the user. Next subsections describe the training steps in details.

3.1 Filtering Tags

The process of filtering noisy tags is divided into two phases. The first one is conducted to amend or discard misspelled, incorrect and/or incomplete terms,

being most of the process supported by lexicon resources and syntactic analysis of the terms. The second phase consists of checking the meaning of the tags in order to decide if such terms are semantically related to the most frequent concepts in the training set. This subsection describes the first phase of the filtering procedure; the second one, because it is executed as a further step of the annotation algorithm, is depicted in details in Subsection 4.3.

The synctatic analysis of the given tags is similar to the one proposed by Cantador et al. [3]. Figure 2 illustrates the main steps. It starts with a lexical checking (Figure 2 (a)) which will discard all stop-words and terms that contain digits and/or their length is less than 2 or bigger than 25 characters.

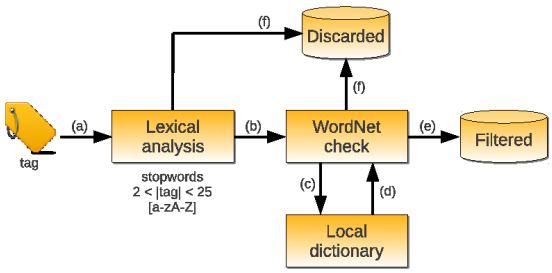

Fig. 2. Filtering procedure.

After that, the tags are checked against WordNet (Figure 2 (b)) in order to find existent words. If it was not found, the tag may be misspelled. Consequently, we check it against a local dictionary (Figure 2 (c)) to find possible correct lexical variations of that term. As opposite to the algorithm proposed by Cantador et al., which uses a Google connector to check lexical variations, we use the Levenshtein distance [11] among the tag and possible candidates, working similarly to *Google Did You Mean*, though the latter also considers proper names. The reason for not having used the Google version is that it restricts the number of queries per day. In our case, tags consisting of proper names will be investigated in future work through the use of Wikipedia ontologies, such as Yago2[5] or DBpedia.

After the local dictionary module has indicated possible lexical variations of the term, they are checked again on WordNet in order to make sure it is a valid English term existent in such database (Figure 2 (d)). If it was found, the term will be considered in the system (Figure 2 (e)); otherwise, it will be discarded (Figure 2 (f)).

3.2 Statistics Calculation

After syntactic analysis, the training images and their associated tags are used to create three statistic tables, as illustrated in Figure 1 (c), (d) and (e). The

[5] http://www.mpi-inf.mpg.de/yago-naga/yago

first two are the tag and concept folksonomies, and the last is the one which relates concepts and tags. As will be described in next section, they are used to infer related tags and concepts from the corpus, supporting the semantic classification.

The folksonomies are based on the relatedness measure described by Cattuto et al. [4], who create a folksonomy based on co-occurrence of tags. They define it as a weighted undirected graph whose set of vertex is the set T of all tags from all images. Two tags t_1 and t_2 are connected by an edge if and only if $t_1, t_2 \in T_s$, where T_s is the set of tags assigned to image $s \in S$. The weight of this edge is given by the number of times t_1 and t_2 co-occurred, i.e., $w(t_1, t_2) = |\{s \in S \mid t_1, t_2 \in T_s\}|$.

For the concept folksonomy, the same process is adopted, but instead of considering co-occurrence of tags, we check the co-occurrence of concepts assigned to each image from the training set. Two concepts c_1 and c_2 are connected by an edge if and only if $c_1, c_2 \in C_s$, where C_s is the set of concepts assigned to image $s \in S$ during training. The weight of this edge is given by the number of times c_1 and c_2 co-occurred, i.e., $w(c_1, c_2) = |\{s \in S \mid c_1, c_2 \in C_s\}|$.

Another statistic table created from the training set is the one which relates concepts and tags. We calculate the number of times a concept and a tag co-occurred in an image. Given a pair (c, t), where $c \in C$ is a concept and C is the set of all concepts, we assume they are related if and only if $c \in C_s$ and $t \in T_s$. A weight between them is also defined, i.e., $w(c, t) = |\{s \in S \mid c \in C_s \ \& \ t \in T_s\}|$.

4 Automatic Annotation

This section describes the automatic annotation algorithm proposed in this paper. It consists of analyzing the available tags of each image in order to infer possible related concepts according to the folksonomies and statistical data constructed during training.

Figure 3 illustrates the overall schema. User tags associated to an image (Figure 3 (a)) are first filtered using the same procedure described in Subsection 3.1 (Figure 3 (b)). After that, some images may have not enough terms to support correct concept classification. Thus, we add a tag enrichment phase (Figure 3 (c)) which will expand the set of available tags with related ones obtained from the tag folksonomy created previously during training.

When an image has enough terms, it can be annotated. This task is accomplished by means of three strategies, illustrated in Figure 3: direct recognition (d), most frequent concepts by tag (e) and post processing (f). As a result of all three classification steps, the multimedia content is annotated with all obtained concepts (Figure 3 (g)). Next subsections describe in details all involved steps.

4.1 Tag Enrichment

Although concepts are inferred based on the content of each tag, it is possible to improve the classification results by enriching the set of tags with related terms

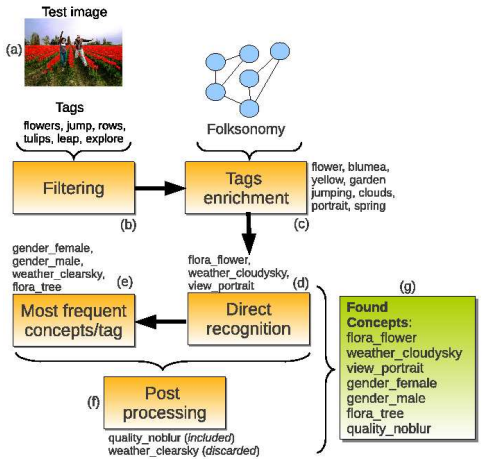

Fig. 3. Annotation procedure.

that users usually use together to tag a resource. When more tags are available, the classifier will be able to better predict concepts because more information about the content will be analyzed.

The tag enrichment procedure is implemented in our system by considering a number of related terms t' to a given tag t. This relatedness between tags is defined proportionally to the edge weights of the tag folksonomy; consequently, given a tag $t \in T$, the tags that are most related to it are all the tags $t' \in T$ with $t \neq t'$ such that $w(t, t')$ is maximal.

Let us consider, for instance, the image illustrated in Figure 4, whose set of tags are "pittsburgh", "sky" and "clouds". Applying the tag enrichment procedure will result in the three subsets illustrated in Figure 4 (a), where the terms that are most related to "pittsburgh" are "sky", "clouds", "yellow", "downtown" and "tree"; the same for the original tags "sky" and "clouds".

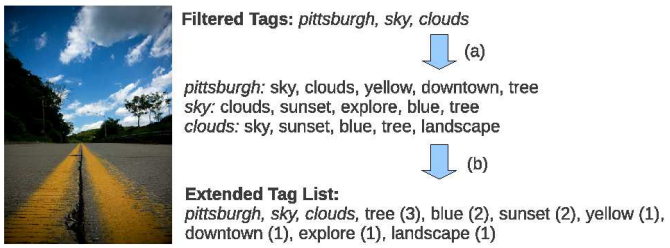

Fig. 4. Tag enrichment procedure.

As a result of gathering the related terms from all tags $t \in T_s$, we obtain a list R_s composed of related terms and associated weights. A weight value is calculated according to the corresponding term's frequency of retrieval by using all tags t. Then, these weights are used to sort the list in descending order. Figure 4 (b) illustrates this step, which results in a list R_s composed of the most frequent terms occurred in the previous subsets.

Following, we empirically define a minimal number of tags $\bar{T}$ to allow each image to be classified. The value of $|\bar{T}|$ includes the terms already associated by the user, plus the related terms obtained from the tag folksonomy. Thus, we consider the n_s first terms from list R_s, whose value is given by:

$$n_s = \begin{cases} |\bar{T}| - |T_s| & \text{if } |T_s| < |\bar{T}| \\ 0 & \text{otherwise.} \end{cases}, \quad \text{or} \tag{1}$$

Consequently, if we set $|\bar{T}| = 5$, the image illustrated in Figure 4 will be classified using the original tags "pittsburgh", "sky" and "clouds", plus the extended ones "tree" and "blue".

4.2 Direct Recognition

The annotation task effectively starts with direct recognition (Figure 3 (d)), which will check whether a user has tagged an image with a term that explicitly refers to a concept. In order to accomplish that, each predefined concept was manually designated a meaningful alias, which was chosen to be the most likely term used by an individual to tag a resource with that concept. For instance, for concept "flora_flower", its alias was defined "flower"; for concept "weather_rainbow", its alias was defined "rainbow". Furthermore, with the objective to improve the results of direct recognition, we also consider the fact that many tags may be in the plural or singular forms, and users may use a synonym word to refer to the same concept. Thus, we first use an inflector object to transform all tags and concepts to their singular form, in addition to a synonym vocabulary that contains different synonyms for each concept. Consequently, if a user has tagged an image with "kids", so the system will first transform it to the singular form "kid", which is a synonym of "child". As a result, as "child" is the alias for the concept "age_child", this concept will be added into the annotation of the given image.

4.3 Most Frequent Concepts

The next step of concept annotation is to obtain the most frequent concepts by tag using the concept-tag relation described in Subsection 3.2. Given a tag $t \in T_s$, we collect the C_t most frequent concepts related to it, i.e., C_t will have a subset of concepts $c \in C$ such that $w(c, t)$ is maximal. Considering the same example presented in Figure 4, in Figure 5 (a) we illustrate this step by listing the five most frequent concepts for each given tag: "pittsburgh", "sky", "clouds", "tree" and "blue".

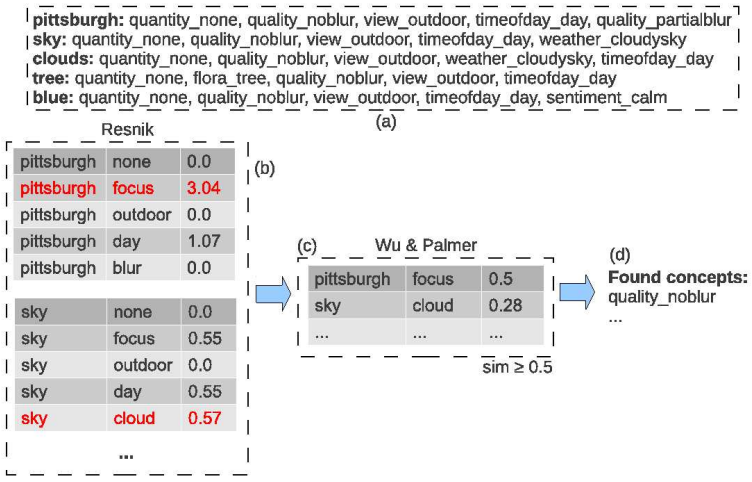

Fig. 5. Most frequent concepts classifier.

Once we have some concepts candidates given by C_t, it is still necessary to check if they are semantically related to the tag t. This is because the concept-tag relation only provides concepts relying on statistic information, it does not ensure that the most frequent concepts have a semantic relation to the given tag. For instance, the tag "explore" is frequently used by users to tag resources in many contexts, though it is not semantically related to the image's actual concepts. Consequently, if this tag is used to retrieve the most frequent concepts, the list will be prone to noise. To solve this problem, we use two semantic similarity measures to decide whether $c' \in C_t$ will be adopted in final annotation or not. The first measure, sim_{wp}, is the Wu & Palmer approach [19], which is based on the topology of the WordNet ontology. The second one, sim_{res}, is the Resnik semantic similarity measure [17], which is based on information content.

Therefore, our algorithm will combine both similarity measures in order to find a concept $c' \in C_t$ that is semantically related to tag t, that is:

$$c_s+ = \begin{cases} c' & \text{if } sim_{wp}(\max_{c' \in C_t}[sim_{res}(alias(c'), t)], t) \geq \Gamma \\ \oslash & \text{otherwise} \end{cases}, \tag{2}$$

where Γ is a threshold, and $alias(c')$ returns the alias of concet c' as explained in Subsection 4.2. This process is executed for all tags associated to image s. Thus, the system will find a set of concepts C_s such that each element $c_s \in C_s$ is a concept that is most semantically similar to a tag $t \in T_s$.

Considering the previous example, Figure 5 (b) illustrates the application of Resnik semantic similarity between each tag and the alias of each concept candidate. For each tag, we select the concept with highest similarity, which is computed again using the Wu & Palmer metric (Figure 5 (c)). Then, using a

predefined value for Γ, we select the concepts to be included in the annotation of the given image (Figure 5 (d)).

4.4 Post Processing

As the final step of the annotation task, we designed a post processing procedure whose goals are twofold: i) to include related concepts to the ones already found; and ii) to discard mutual exclusive concepts by comparing each pair of annotated concepts using a predefined set of heuristics.

In the first case, we use a Naïve Bayes approach [11] in order to determine whether, given an annotated concept c, other concepts should be included as well. Thus, we define:

$$c_s+ = \begin{cases} c_j \in C & \text{if } P(c_j|c) \geq P(!c_j|c) \\ \oslash & \text{otherwise} \end{cases} \quad , \tag{3}$$

where $P(c_j|c)$ is the probability of including c_j in the annotation given the annotated concept c; and $P(!c_j|c)$ is the probability of **not** including c_j in the annotation given the annotated concept c. We estimate both probabilities as:

$$P(c_j|c) \approx P(c_j)P(c|c_j) = \frac{|S_{c_j}|}{|S|} \frac{w(c_j, c)}{\sum_{c_i \in C} w(c_i, c_j)} \quad , \tag{4}$$

$$P(!c_j|c) \approx P(!c_j)P(c|!c_j) = \frac{|S_{!c_j}|}{|S|} \frac{w(!c_j, c)}{\sum_{c_i \in C} w(c_i, !c_j)} \quad , \tag{5}$$

where S_{c_j} is the subset of images from the training set annotated with concept c_j; and $S_{!c_j}$ is the subset of images from the training set which were **not** annotated with concept c_j.

After including related concepts, the final procedure of our annotation tool is to discard mutual exclusive concepts. To do that, we first design manually a list of pairs of concepts which are mutual exclusive, such as "view_indoor" versus "view_outdoor", "sentiment_active" versus "sentiment_inactive", and so on. Then, given a pair of conflicting concepts (c_1, c_2) in the preliminary annotation list, we apply the following heuristics in order to discard the concept(s) in conflict:

1. If c_1 and c_2 were found by direct recognition, then keep both concepts.
2. If c_1 was found by direct recognition and c_2 was not, then discard c_2.
3. If c_1 and c_2 were not found by direct recognition, discard the concept with less votes from all classifiers.
4. If c_1 and c_2 were not found by direct recognition and both have the same amount of votes, discard both concepts.

5 Results

This section presents the official results obtained by our group in the ImageCLEF 2012 Photo Annotation Task [18]. Table 1 presents the overall results. We submitted two runs to be evaluated: the "proposed" one, which is the implementation of the techniques described in this paper; and a simpler version based only on tags enrichment and Naïve Bayes. This year, ten groups submitted at least one run in the textual category; comparing the best results from each group, our technique was able to achieve the eighth place.

Table 1. Overall Results.

			By Photo		By Concept	
Classifier	MiAP	GMiAP	Micro-F1	Macro-F1	Micro-F1	Macro-F1
Naïve Bayes	0.1521	0.0894	0.3532	0.3877	0.3532	0.2152
Proposed	0.1724	0.1140	0.3389	0.3460	0.3389	0.2586

Figure 6 presents the F1 results for each of the 93 predefined concepts. We note that better results were achieved for the concepts in the classes "fauna" and "transport", once users tend to use the concepts names to tag their resources. Furthermore, other particular concepts were also able to be correctly classified, such as "combustion_fireworks", "setting_fooddrink", "quantity_none" and "quality_noblur", being the first two recognized mostly by direct classification and most frequent concepts by tag, and the last two by the post processing step.

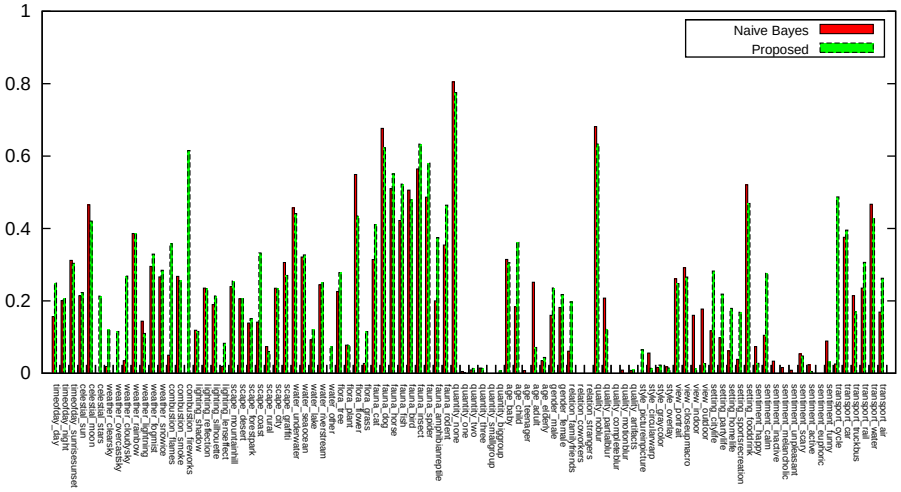

Fig. 6. Results by concept.

6 Final Remarks

This paper presented the results of the automatic annotation tool evaluated by ImageCLEF 2012 Photo Annotation Task. It uses different strategies based on co-occurrence of tags and concepts in order to reduce two problems related to the use of tags in classification: noise and incompleteness. The first one is supported by a multi-phase filtering procedure that amends misspelled tags and considers only those semantically related to the actual context of the image. The second one, in turn, is reduced with a tag enrichment activity which aggregates related terms to the actual list of tags in order to make annotation more effective.

The results shown by this evaluation indicate that more research should be made in order to improve the quality of annotations. This shall be made in future work, as we plan to investigate how to recognize specific concepts which are difficult to be found using only textual features.

7 Acknowledgments

The authors would like to thank the financial support from FAPESP, process number 2011/17366-2.

References

1. K. Amel, A. Benammar, and C. B. Amar. REGIMvid at ImageCLEF2011: Integrating Contextual Information to Enhance Photo Annotation and Concept-based Retrieval. In V. Petras, P. Forner, and P. D. Clough, editors, *Working Notes of CLEF 2011*, Amsterdam, The Netherlands, 2011.
2. D. Brezeale and D. J. Cook. Automatic Video Classification: A Survey of the Literature. *IEEE Transactions on Systems, Man, and Cybernetics*, 38(3):416–430, 2007.
3. I. Cantador, M. Szomszor, H. Alani, M. Fernández, and P. Castells. Enriching Ontological User Profiles with Tagging History for Multi-Domain Recommendations. In *1st International Workshop on Collective Semantics: Collective Intelligence & the Semantic Web (CISWeb 2008)*, Tenerife, Spain, 2008.
4. C. Cattuto, D. Benz, A. Hotho, and G. Stumme. Semantic Grounding of Tag Relatedness in Social Bookmarking Systems. In A. P. Sheth, S. Staab, M. Dean, M. Paolucci, D. Maynard, T. W. Finin, and K. Thirunarayan, editors, *The Semantic Web – ISWC 2008*, volume 5318 of *Lecture Notes in Computer Science*, pages 615–631, Berlin/Heidelberg, 2008. Springer.
5. R. L. Cilibrasi and P. M. B. Vitanyi. The google similarity distance. *IEEE Trans. on Knowl. and Data Eng.*, 19(3):370–383, 2007.
6. B. Daróczy, R. Pethes, and A. A. Benczúr. SZTAKI @ ImageCLEF 2011. In V. Petras, P. Forner, and P. D. Clough, editors, *Working Notes of CLEF 2011*, Amsterdam, The Netherlands, 2011.
7. R. Izawa, N. Motohashi, and T. Takagi. Annotation and Retrieval System Using Confabulation Model for ImageCLEF2011 Photo Annotation. In V. Petras, P. Forner, and P. D. Clough, editors, *Working Notes of CLEF 2011*, Amsterdam, The Netherlands, 2011.

8. D. D. Lewis, Y. Yang, T. G. Rose, and F. Li. RCV1: A New Benchmark Collection for Text Categorization Research. *J. Mach. Learn. Res.*, 5:361–397, 2004.
9. W. Li, J. Min, and G. J. F. Jones. A Text-Based Approach to the ImageCLEF 2010 Photo Annotation Task. In M. Braschler, D. Harman, and E. Pianta, editors, *Working Notes of CLEF 2010*, Padua, Italy, 2010.
10. N. Liu, Y. Zhang, E. DellandrÃ©a, S. Bres, and L. Chen. LIRIS-Imagine at ImageCLEF 2011 Photo Annotation Task. In V. Petras, P. Forner, and P. D. Clough, editors, *Working Notes of CLEF 2011*, Amsterdam, The Netherlands, 2011.
11. C. D. Manning, P. Raghavan, and H. Schtze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.
12. K. Nagel, S. Nowak, U. Kuhhirt, and K. Wolter. The Fraunhofer IDMT at ImageCLEF 2011 Photo Annotation Task. In V. Petras, P. Forner, and P. D. Clough, editors, *Working Notes of CLEF 2011*, Amsterdam, The Netherlands, 2011.
13. S. N. Patel and G. D. Abowd. The ContextCam: Automated Point of Capture Video Annotation. In F. Khendek and R. Dssouli, editors, *UbiComp 2004: Ubiquitous Computing*, volume 3205 of *Lecture Notes in Computer Science*, pages 301–318. Springer, 2004.
14. J. Read, B. Pfahringer, G. Holmes, and E. Frank. Classifier chains for multi-label classification. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases: Part II*, ECML PKDD '09, pages 254–269, Berlin, Heidelberg, 2009. Springer-Verlag.
15. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.
16. E. Spyromitros-Xioufis, K. Sechidis, G. Tsoumakas, and I. Vlahavas. MLKD's Participation at the CLEF 2011 Photo Annotation and Concept-Based Retrieval Tasks. In V. Petras, P. Forner, and P. D. Clough, editors, *Working Notes of CLEF 2011*, Amsterdam, The Netherlands, 2011.
17. P. R. Sun. Using information content to evaluate semantic similarity in a taxonomy. In *In Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 448–453, 1995.
18. B. Thomee and A. Popescu. Overview of the ImageCLEF 2012 Flickr Photo Annotation and Retrieval Task. In *CLEF 2012 working notes*, Rome, Italy, 2012.
19. Z. Wu and M. Palmer. Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pages 133–138, Stroudsburg, PA, USA, 1994. Association for Computational Linguistics.
20. S. Zhu, G. Wang, C.-W. Ngo, and Y.-G. Jiang. On the sampling of web images for learning visual concept classifiers. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, CIVR '10, pages 50–57, New York, USA, 2010.
21. A. Znaidia and H. Le Borgne. CEA LISTs participation to Visual Concept Detection Task of ImageCLEF 2011. In V. Petras, P. Forner, and P. D. Clough, editors, *Working Notes of CLEF 2011*, Amsterdam, The Netherlands, 2011.